



**CONSTRUCTING POLYMORPHIC VIRUS ANALYSIS SYSTEM
USING BEHAVIOR DETECTION APPROACH**

FAUZI ADI RAFRASTARA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS THESIS*

JUDUL: CONSTRUCTING POLYMORPHIC VIRUS ANALYSIS SYSTEM USING BEHAVIOR DETECTION APPROACH

SESI PENGAJIAN: _____

Saya FAUZI ADI RAFASTARA

(HURUF BESAR)

Mengaku membenarkan tesis Sarjana ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

_____ SULIT

(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

_____ TERHAD

(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

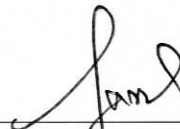
_____ TIDAK TERHAD



(TANDA TANGAN PENULIS)

Alamat Tetap: Jln. Taman Borobudur II
Rt. 09 R.W. 10 kel. Kembang arum,
Semarang - Central Java - Indonesia

Tarikh: 29 NOV 2010



(TANDA TANGAN PENYELIA)

DR. MOHD. FAIZAL BIN ABDULLAH

Nama Penyelia

Tarikh: 29 NOV 2010

CATATAN: * Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM).

** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

**CONSTRUCTING POLYMORPHIC VIRUS ANALYSIS SYSTEM
USING BEHAVIOR DETECTION APPROACH**

FAUZI ADI RAFRASTARA

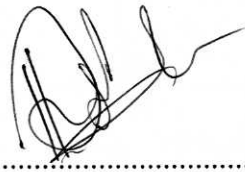
**This report is submitted in partial fulfillment of the requirements for the
Master of Computer Science (Software Engineering and Intelligence)**

**FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA
2010**

DECLARATION

I hereby declare that this project report entitled “Constructing Polymorphic Virus Analysis System Using Behavior Detection Approach” is the result of my own research except as cited in the references. The report has not been accepted for any degree and is not concurrently submitted in candidature of any other degrees.

Signature



:

Name

: FAUZI ADI RAFASTARA

Date

: 29 NOV 2010

APPROVAL

I hereby declare that I have read through this project report and in my opinion this project report is sufficient in terms of scope and quality for the award of the degree of Master of Computer Science (Software Engineering and Intelligence).

Signature

: 

Supervisor

: DR. MOHD. FAIZAL BIN ABDOLLAH

Date

: 29 Nov 2010

DEDICATION

To my family for their love, caring, sacrifice, and support during the process of achieving this milestone in my life.

ACKNOWLEDGMENT

Alhamdulillah, all praises to Allah, for the strengths and His blessing in completing this project entitled: *Constructing Polymorphic Virus Analysis System, Using Behavior Detection Approach*.

Special appreciation goes to my supervisor, Dr. Mohd. Faizal bin Abdollah, for his supervision and constant support. His invaluable help of constructive comments and suggestions throughout the experimental and project works have contributed to the success of this research.

I would like to express my appreciation to the Dean, Faculty of Information and Communication Technology, Prof. Dr. Shahrin bin Sahib @ Sahibuddin and also to the Deputy Dean, Faculty of Information and Communication Technology, Prof. Nanna Suryana Herman and Prof. Madya Dr. Burairah Hussin, for their support and help towards my postgraduate affairs. I also would like to thank Dr. I Gedhe Pramudya Ananta, for his very important suggestions in the preparation of this thesis. My acknowledgement goes to all the technicians and office staffs of Faculty of Information and Communication Technology as well, especially for En. Badrolhisam and En. Ridzuan for their co-operations.

I am also very thankful to Prof. Dr. Nanna Suryana Herman and his wife, who have been pleasing to be my foster parent during my studies in Malaysia. Without their continued support and interest, this project would not have been same as presented here.

I am also indebted to Dian Nuswantoro University (UDINUS) for their kindness in giving financial aid, during my study here. Special thanks also go to Dr. Ir. Edi Noersasongko, Dr. Abdul Syukur, Dr. Kusni Ingsih, Dr. St. Dwiarto Utomo, Mr. A. Zainul Fanani and Mrs. Yunita, for their tremendous support.

Sincere thanks to all my friends and senior, especially Mr. Affandy, Abang Fahmi, Mas Ricardus, Mas Yusuf, Indi, Nikita, Mb. Dhani, En. Zul, Mas Catur, Ferda, Mas Andi Juliarno, Rabeea, Sa'ad, Ashraf Omoush, Mahir, Kak Ayu, Kak Aznor, Syazwani, UDINUS Gang '10, PPI UTeM, UTeM Tennis Club and others for their kindness and moral support during my study. Thanks for the friendship and memories.

Last but not least, my deepest gratitude goes to my beloved parents; Mr. Mohd. Yazid Jamil and Mrs. Siti Johariyah and also to my sisters (Tika & Atya) and brother (Reza) for their endless love, prayers and encouragement. To those who indirectly contributed in this research, your kindness means a lot to me. Thank you very much.

TABLE OF CONTENTS

	PAGE
DECLARATION	ii
APPROVAL	iii
DEDICATION	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	xi
LIST OF APPENDICES	xiii
ABSTRACT	xiv
ABSTRAK	xv
CHAPTER	
1. INTRODUCTION	1
1.1 Introduction	1
1.2 Background of Study	2
1.3 Problem Statement	3
1.4 Research Questions	3
1.5 Research Objectives	4
1.6 Research Methodology	4
1.7 Research Scope	6
1.8 Research Contribution	7
1.9 Project Report Overview	7
1.10 Conclusion	8
2. LITERATURE REVIEW	10
2.1 Introduction	10
2.2 Computer Virus	11
2.2.1 Traditional Virus	12
2.2.2 Polymorphic Virus	13
2.3 MD5 Checksum	15
2.4 Current Virus Analysis System	18
2.4.1 CWSandbox	18
2.4.2 Capture	19
2.4.3 MBMAS	21

2.4.4	Joebox	23
2.4.5	ThreatExpert	25
2.5	Summary	26
3.	RESEARCH METHODOLOGY	27
3.1	Introduction	27
3.2	Research Phase	28
3.2.1	Phase 1 – Study of Computer Virus and Anti Virus (AV)	29
3.2.2	Phase 2 – Study and Analysis of the Literature Review	29
3.2.3	Phase 3 – Collecting Data	29
3.2.4	Phase 4 – Implementation	30
3.2.5	Phase 5 - Testing	30
3.2.6	Phase 6 – Result Analysis	31
3.2.7	Phase 7 – Conclusion	31
3.3	System Development Methodology	31
3.3.1	Analysis	32
3.3.2	Design	33
3.3.3	Coding	33
3.3.4	Testing	34
3.4	Data Collection	34
3.5	Summary	39
4.	IMPLEMENTATION	40
4.1	Analysis	40
4.1.1	Problem Analysis	41
4.1.1.1	The Current System	41
4.1.1.2	Problem Identification	43
4.1.1.3	The Proposed Architecture	44
4.1.2.	System Requirement Analysis	45
4.1.2.1	Hardware Requirement	46
4.1.2.2	Software Requirement	47
4.1.2.3	User Requirement	47
4.2	Design	48
4.2.1	Context Diagram	48
4.2.2	Data Flow Diagram Level 0	50
4.2.3	Architecture	52
4.2.4	Input Output Design	54
4.2.4.1	Input Design	54
4.2.4.2	Output Design	88
4.3	Coding	89
4.3.1	Virus Behavior Monitoring Tool (VBMT)	89
4.3.1.1	Pre-Monitoring (Preparation)	90
4.3.1.2	Monitoring Process	92
4.3.1.3	Post-Monitoring	96
4.3.2	Virus Behavior Analysis Tool (VBAT)	97
4.4	Summary	100

5. TESTING	101
5.1 Testing Process	101
5.2 Summary	116
6. SUMMARY AND CONCLUSION	117
6.1 Summary	117
6.2. Conclusion	118
6.3 Future Work	119
REFERENCES	120
APPENDIX A	126
APPENDIX B	134
APPENDIX C	138
APPENDIX D	141

LIST OF TABLES

TABLE	TITLE	PAGE
3.1	List of viruses which tested in data collection phase	37
3.2	List of locations of virus attack in registry	38
3.3	List of locations of virus attack in Windows	38
4.1	Hardware Requirement	46
4.2	Software Requirement	47
4.3	Description of Textbox1 Control in the first form	57
4.4	Description of Textbox1 control in the first form	58
4.5	Description of Button1 control in the first form	59
4.6	Description of OptionButton1 control in the first form	60
4.7	Description of SpinnerBox1 control in the first form	61
4.8	Description of OptionButton2 control in the first form	62
4.9	Description of Button2 control in the first form	63
4.10	Description of Button3 control in the first form	64
4.11	Description of SSTab1 control in the third form	66
4.12	Description of Textbox2 control in the third form	67
4.13	Description of Button4 control in the third form	68
4.14	Description of Textbox3 control in the third form	69
4.15	Description of Button5 control in the third form	70
4.16	Description of Button6 control in the third form	71
4.17	Description of Button3 control in the third form	72
4.18	Description of ListView1 control in the fourth form	74
4.19	Description of TextBox4 control in the fourth form	75
4.20	Description of TextBox5 control in the fourth form	76
4.21	Description of ListView1 control in the fourth form	77

4.22	Description of ListView3 control in the fourth form	78
4.23	Description of ListView3 control in the fourth form	79
4.24	Description of Button8 control in the fourth form	80
4.25	Description of Button9 control in the fourth form	81
4.26	Description of Button10 control in the fourth form	82
4.27	Description of Button3 control in the fourth form	83
4.28	Description of ListView5 control in the fifth form	85
4.29	Description of TextBox6 control in the fifth form	86
4.30	Description of Button12 control in the fifth form	87
5.1	Comparison between preliminary research result and testing result	115

LIST OF FIGURES

FIGURE	TITLE	PAGE
1.1	Research Methodology	5
2.1	Instances of encrypted and decrypted polymorphic virus bodies (Szor, 2005)	14
2.2	Message digest generation using MD5 (Stallings, 2003)	17
2.3	Capture architecture diagram	20
2.4	Architecture of MBMAS (FuYong et al., 2009)	22
2.5	Architecture of Joebox (Joebox, 2010)	23
3.1	Research Phases	28
3.2	Waterfall Model (Saleh, 2009)	32
4.1	Context Diagram of the proposed system	49
4.2	DFD Level 0 of the proposed system	51
4.3	Architecture of VMAS+	53
4.4	Architecture of VMAS+ in the virtual environment	54
4.5	Design of main form	56
4.6	Form design to browse a file	65
4.7	Design of VBAT form	65
4.8	Design of Monitoring Form	73
4.9	Design of Analysis Form	84
4.10	VMAS+ main window	91
4.11	Browse File window	92
4.12	Submitting the virus file and adjusting the timeout limit	92
4.13	Window <i>Monitoring Process</i>	93
4.14	Pseudo code of file monitoring	94
4.15	Pseudo code of registry monitoring	95

4.15	Pseudo code of process monitoring	96
4.16	VBAT window	98
4.17	Result Analysis window	99
4.18	Pseudo code of VBAT	100
5.1a	Flow Chart of Testing Phase	102
5.1b	Flow Chart of Testing Phase (Continued)	103
5.2	Monitoring process of <i>W32.HLLW.Benfgame.B (Fasong)</i> virus on the first PC	105
5.3	Monitoring process of <i>W32.HLLW.Benfgame.B (Fasong)</i> virus on the second PC	105
5.4	Conclusion of result analysis against <i>W32.HLLW.Benfgame.B (Fasong)</i>	106
5.5	Monitoring process of <i>W32.HLLW.Lovgate.J@mm</i> virus on the first PC	107
5.6	Monitoring process of <i>W32.HLLW.Lovgate.J@mm</i> virus on the second PC	108
5.7	Conclusion of result analysis against <i>W32.HLLW.Lovgate.J@mm</i>	108
5.8	Monitoring process of <i>W32.SillyFDC (Brontok)</i> virus on the first PC	109
5.9	Monitoring process of <i>W32.SillyFDC (Brontok)</i> virus on the second PC	110
5.10	Conclusion of result analysis against <i>W32.SillyFDC (Brontok)</i>	110
5.11	Monitoring process of <i>W32.Klez.E@mm</i> virus on the first PC	111
5.12	Monitoring process of <i>W32.Klez.E@mm</i> virus on the second PC	112
5.13	Conclusion of result analysis against <i>W32.Klez.E@mm</i> virus	113
5.14	Monitoring process of <i>W32.Virut.CF</i> virus on the first PC	113
5.15	Monitoring process of <i>W32.Virut.CF</i> virus on the second PC	114
5.16	Conclusion of result analysis against <i>W32.Klez.E@mm</i> virus	114

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	LIST OF VIRUSES' BEHAVIOR	126
B	LIST OF NEW FILES CREATED BY VIRUSES	134
C	CHECKSUMS COMPARISON	138
D	RELATED PUBLICATIONS	141

ABSTRACT

The current antivirus products were only able to detect the existence of viruses, but it could not record the activity or behavior of viruses. Inability of antivirus to record the viruses' behavior made difficult certain users who want to know the behavior of viruses as well as to know the category or classification of certain viruses. Actually, there were several architectures or tools proposed, but they still could not answer the needs of those certain users who want to know the classification of virus that they test.

In this project, we studied the current types of viruses as well as current virus monitoring and analysis system. This study came up with the problems that become basic of this research. Here, we proposed an architecture and a system, which are able to monitor the viruses' behavior and classify those viruses whether as a traditional or polymorphic virus. Preliminary research was conducted to get the current virus behaviors and to find out the certain parameters, which are usually used by viruses to attack the computer target. Finally, we applied "test bed environment" to test our system by releasing several viruses in a real environment, and attempt to capture their behaviors. These activities were followed by generating the conclusion that the tested or monitored virus is classified as a traditional or polymorphic virus.

ABSTRAK

Produk antivirus yang ada pada masa kini hanya mampu untuk mengesan kewujudan virus, namun ia tiada kemampuan untuk mencatatkan aktiviti serta sifat-sifat bagi sesuatu virus. Bagi sesetengah pengguna, ini adalah salah satu kesukaran bagi mereka untuk mengetahui sifat-sifat sesuatu virus termasuk untuk mengetahui kategori bagi satu-satu virus. Sebenarnya, terdapat beberapa alat yang telah dicadangkan oleh beberapa pengkaji, namun persoalan mengenai keperluan bagi sesetengah pengguna yang ingin tahu tentang kategori virus yang telah di uji dan kaji masih belum terjawab.

Kajian yang dilakukan dalam thesis ini adalah satu kajian mengenai jenis-jenis virus dan antivirus masa kini, serta system pemantau dan penganalisis, untuk menjumpai punca masalah yang mana ianya juga merupakan asas kepada kajian ini. Pada projek ini, kami mencadangkan sistem dan senibina yang mampu memantau sifat-sifat virus dan mampu untuk mengkategorikan samada virus tersebut merupakan virus tradisional atau virus polimorfik. Kajian pada peringkat awalnya dijalankan untuk mendapatkan sifat-sifat virus pada masa kini dan untuk mencari parameter tertentu yang biasanya digunakan oleh virus untuk menyerang computer yang disasarkan. Akhirnya, kami menggunakan “test bed environment” untuk menguji sistem kami dengan melepaskan virus dalam persekitaran yang nyata, dan cuba untuk menangkap perilaku dan sifat-sifat mereka, dan diikuti dengan membuat kesimpulan samada virus yang diuji atau dipantau tadi dapat diklasifikasikan sebagai virus tradisional atau polimorfik.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Nowadays, we all live in the digital era, which most information moves from one place to another digitally. The information can be derived easily from everywhere and send it to whoever, only in minutes or even seconds. Unfortunately, wherever we are, including in this digital information era, threats always exist, perhaps in the different shapes. One of the popular threats which always peering us in this era, is Computer Virus.

The virus is a threat, because it can do bad things to whomever. It can make the computer becomes slow, broken, or even it can delete the data. The virus can run automatically and hide the process, so that users cannot see the processes and activities, which are done by virus. What can users see from the virus is what they have done.

1.2 Background of Study

The current tool that nearest to virus world now is AntiVirus (AV), as it can detect and combat almost all kinds of virus. However, the AntiVirus cannot produce the analysis report, which is able to describe the viruses' behavior in details. Analysis report is quite important for those who want to learn how viruses actually act. Furthermore, people can eliminate the viruses from their PC and recover the Operating System from viruses attack by reading the virus behavior analysis report (FuYong, DeYu, & JingLin, 2009). Such kind of report is only provided by several tools, which mostly do not have a capability in virus detection system, such as CWSandbox, Capture, MBMAS, Joebox and ThreatExpert (FuYong, DeYu, & JingLin, 2009).

The aforementioned tools indeed are able to produce the behavior analysis report in details. Unfortunately, by using these tools, the type of malicious file, that have been tested, cannot be known. Even though the analysis report can be derived, it is not easy to determine which virus file is classified as traditional or polymorphic only by reading this report (Bayer, 2005; FuYong, DeYu, & JingLin, 2009).

Based on the explanation above, a polymorphic virus analysis tool, which can solve both problems above, which are able to report the viruses' behavior as well as classify the tested virus, whether it is traditional or polymorphic virus needs to be developed.

1.3 Problem Statement

The problem statement in this report reviews the weaknesses that are found in the real world. There are two main reasons why virus behavior analysis system needs to be developed:

1. The existing of antivirus products cannot impede the curiosity of common users to know the activity or behavior of the viruses. Antivirus is only able to produce the report which inform to us which file is malicious. It cannot produce the analysis report that describes the viruses' behavior. By this case, it could be difficult for common users to learn how viruses acts, and further, they will face the difficulties when they want to combat or eliminate the virus and recover the operating system by themselves.
2. There are several tools and techniques proposed by other researchers which are able to analyze several types of malwares. However, they still can not distinguish and classify which file is classified as a traditional or polymorphic virus.

1.4 Research Questions

The main research questions for this study are given as follows:

1. What is polymorphic virus and how does it act and propagate?
2. How to monitor and analyze the behavior of virus, produce report in details, and classify the virus whether it is traditional or polymorphic virus?
3. How to test and validate the architecture which is used by the virus behavior analysis system to monitor and analyze the behavior of virus, produce report in details, and classify the virus whether as a traditional or polymorphic virus?

1.5 Research Objectives

Based on the research questions above, there are two objectives for this research.

Those objectives are:

1. To study the viruses' characteristic, behavior, and propagation technique which used by traditional and polymorphic virus.
2. To propose an architecture which can be used to monitor and analyze the behavior of virus, produce report in details, and classify the virus whether it is traditional or polymorphic virus.
3. To test and validate the architecture refering to the virus behavior analysis.

1.6 Research Methodology

The research methodology that has been used throughout the research is an iterative process where the study specifications consists of literature review by referring to some documents and books that lead to some ideas. Furthermore, Virus Monitoring and Analysis System (VMAS) for traditional and polymorphic virus will be developed.

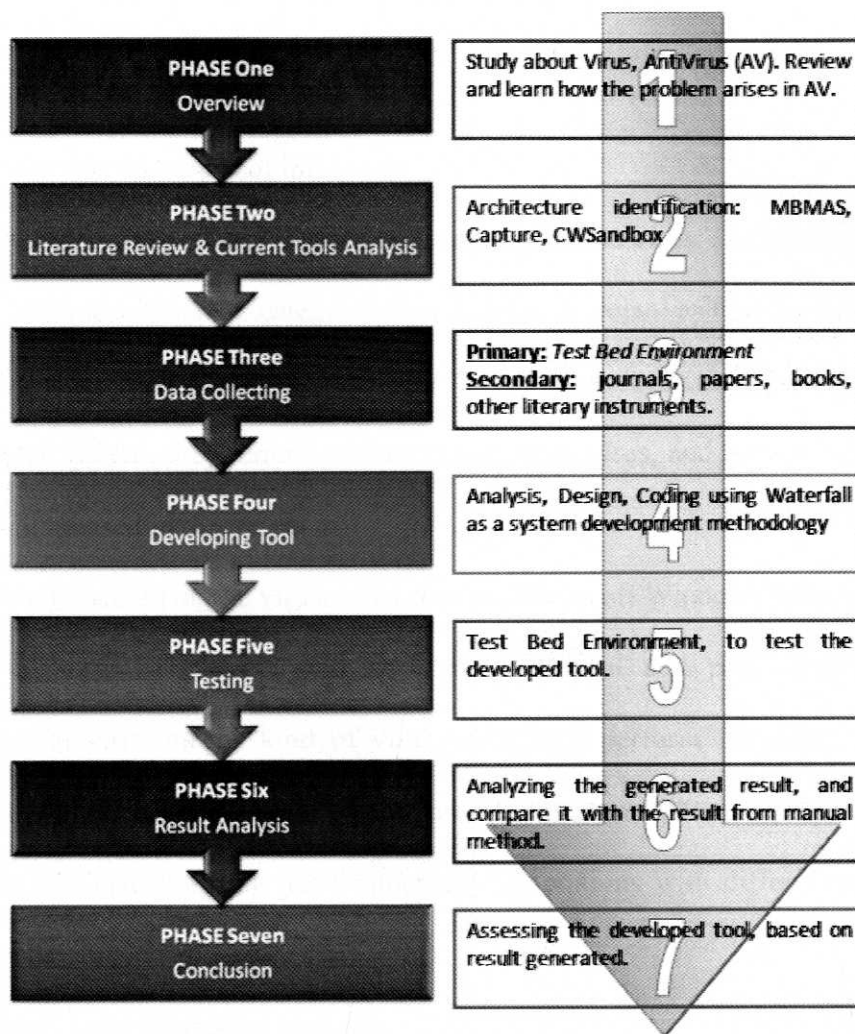


Figure 1.1: Research Methodology

In order to achieve the objectives, this research is divided into seven different phases. Figure 1.1 shows the research phases of this research, which are: research overview, literature review, collecting the required data, developing VMAS tool to monitor and analyze polymorphic and traditional virus, perform testing for the developed tool, and finally assessing the developed tool based on result generated.

1.7 Research Scope

There are several scopes of this study:

1. This study focuses on one type of malwares, which is virus. Eventhough the malware that will be tested here is categorized as a worm or trojan, as long as they have several virus characteristic, such as: perform infection or effecting permanent effect in computer system, they remain will be classified as virus, and these kinds of malwares also will be used in this research.
2. This study focuses on the viruses that runs on Microsoft Windows32-bit platforms.
3. Kind of virus which analyzed here are traditional and polymorphic virus only. Traditional virus means kind of virus when they perform propagation, all of files created will have same signature with parent's signature. Whereas polymorphic virus is a virus that can duplicate itself into many generations with different signatures, but their behaviors are still same.
4. This study only focuses on host side.
5. This study only focuses on file activities, Windows registry activites, and process activities, in which all those activities were done by virus.
6. This study only focuses on virus behavior analysis and distinction between traditional and polymorphic virus by comparing the signature between mother and the offspring.
7. To differentiate the signature among files, this study uses MD5 algorithm in which it has been widely used by antivirus to capture and detect the signature of virus.

1.8 Research Contribution

The contributions of this research are given as follows:

1. This system comes up with the analysis report, and it can be utilized by interested person to get a quick understanding of the purpose of a virus, either traditional or polymorphic virus.
2. By using this system, common users can learn more about virus, and find out which virus is classified as a traditional or polymorphic virus, and they can see the differences between traditional and polymorphic virus, especially in term of signature generated.
3. Beside the proposed architecture can be used by antivirus software directly, the produced report can be used by common users to eliminate the virus, either traditional or polymorphic virus, and recover the operating system as well.
4. With the ability to classify a virus, further, it can be utilized to create intelligent virus remover which is able to clean the system effectively either from traditional or even polymorphic virus attack.

1.9 Project Report Overview

Chapter 1, Introduction, this chapter provides a justification and also background information of this research, problem statement, research question, objective, methodology, scope and research contribution.